MASTER COPY KEEP THIS COPY FOR REPRODUCTION PURPOSES

AD-A263 985 [...]TION PAGE

Form Approved OMB No. 0704-0188 (2)

Pu[...] gat[...] coll[...] Dav[...] [...]rage 1 hour per response, including the time for reviewing instructions, searching existing data sources [...]he collection of information. Send comments regarding this burden estimate or any other aspect of this [...] Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson [...]anagement and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. [...] [...]E 2/26/93

3. REPORT TYPE AND DATES COVERED Final 30 Sep 91 - 31 Dec 92

4. TITLE AND SUBTITLE Intelligent Support for Human Computer Interaction and Decision-making in Distribution planning & Scheduling Systems

5. FUNDING NUMBERS DAAL03-91-G-0328

6. AUTHOR(S)

Drs. Steven F. Roth & Stephen F. Smith

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA. 15213-3890

DTIC ELECTE MAY 07 1993 S D E

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING/MONITORING AGENCY REPORT NUMBER ARO 29354.3-MA

11. SUPPLEMENTARY NOTES

The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

93-09538

13. ABSTRACT (Maximum 200 words)

Research was conducted on two areas related to computer support for decision-making and analysis in complex domins, with a focus on distribution and transportation planning in the Army. The work addressed frameworks and tools for human-computer interaction in systems involving large amounts of diverse information and development of decision-making models.

Research on human-computer interaction involved:

- dynamic display generation for information with more complex structure than previously addressed, including encoding new knowledge of graphic design for this data and more complex graphic styles, and
- design and integration of new interactive data exploration and analysis tools

Research on generation of heuristic models for logistics support included:

- using genetic algorithm (GA) based machine learning for generation of predictive, decision-support models, focusing on scalability and predictive model management in changing environments for ammunition requirements forecasting, and
- performing comparitive experimental tests of techniques which combine conventional statistical approaches with GA approaches to generate forecasting models.

14. SUBJECT TERMS

15. NUMBER OF PAGES 11

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. [...]) Prescribed by ANSI Std. Z39-18 298-102

Intelligent Support for Human Computer Interaction and Decision Making in Distribution Planning & Scheduling Systems

Final Report

Dr. Steven F. Roth and Stephen F. Smith

February 28, 1993

U.S. Army Research Office
Contract Number DAAL03-91-G-0328

Carnegie-Mellon University
5000 Forbes Avenue
Pittsburgh PA 15213

Abstract

Research was conducted on two areas related to computer support for decision-making and analysis in complex domins, with a focus on distribution and transportation planning in the Army. The work addressed frameworks and tools for human-computer interaction in systems involving large amounts of diverse information and development of decision-making models.

Research on human-computer interaction involved:

- dynamic display generation for information with more complex structure than previously addressed, including encoding new knowledge of graphic design for this data and more complex graphic styles, and
- design and integration of new interactive data exploration and analysis tools

Research on generation of heuristic models for logistics support included:

- using genetic algorithm (GA) based machine learning for generation of predictive, decision-support models, focusing on scalability and predictive model management in changing environments for ammunition requirements forecasting, and
- performing comparitive experimental tests of techniques which combine conventional statistical approaches with GA approaches to generate forecasting models.

Problems Studied & Summary of Important Results

Automatic Graphic Presentation:

There have been several goals of our research during the course of this project. Our long term goal is to develop environments in which people can explore and analyze large amounts of diverse information. These environments must have usable mechanisms for creating effective visual representations of data which support the tasks they are performing. Our approach to this problem has been to integrate knowledge-based techniques for graphic design with interactive techniques for exploring and manipulating quantitative and relational data. As we discussed in previous reviews, there were several basic research problems that needed to be addressed.

First, although progress had been made in automating many aspects of display creation, previous work had addressed the automatic selection and composition of graphical techniques for which there was a one-to-one correspondence between data objects and graphical objects. Specifically, this work involved the mapping of data objects and their attributes to graphical objects and their graphical properties. As a result, this work focused exclusively on binary relations (those for which a single data object was related to a value via a simple attribute).

As a result of our experiences in this project with a corpus of data we constructed from Army logistics applications, it was clear that we needed to develop an approach for data structures that are more complex than the binary relations previously represented. Specifically, we needed to develop approaches to the design of displays for N-ary relations. In relational database terms, a relation defines a mapping among domains of values (where N expresses the number of domains). Binary relations (i.e. N is two) are typical in object-attribute-value representations and the only form supported in previous research. Examples from a military transportation database are (with the relation/attribute expressed first):

(Early-Arrival-Date Move-Requirement Date), where a single fact might be:

(EAD A001 C003), expressing the fact that requirement A001 has an EAD of Day C003

(Departing-from Airport move-requirement port), where a single fact might be:
(Departing-from A001 NYC)

(CAPACITY port tons-per-day), where a single fact might be:
(CAPACITY NYC 100,000)

Relations involving more than two domains usually describes a property of an object which varies over time, space, or along any other dimension. Examples include:

(MOVEMENT move-requirement tons-of-cargo date), where a single fact might be:
(MOVEMENT A001 1000 C003), expressing the fact that a movement of a portion (1000 tons) of the cargo specified in the database for unit A001 was moved on day C003. In this case, there would be multiple entries for unit A001, each describing a shipment of a different quantity of its cargo.

(OFF-LOAD ship tons-of-cargo port date move-requirement), where a single fact might be:
(OFF-LOAD USS-HEAP 5,000 Saudi Arabia C005 A001), expressing the fact that 1000 tons of A001's cargo was delivered to Saudi Arabia by the ship USS-HEAP on day C005. Again, there would be multiple occurrences of each of the objects in this fact throughout the database and therefore a one-to-one correspondence between a data object and graphical object is not possible. In this database, there were may records referring to each unit, port, ship, etc.

Relations like these increase the complexity of the graphics design task because they require multiple graphical techniques to express one fact. In contrast, each binary relation could be expressed by a single technique. For example, the Early-Arrival-Date relation could be expressed by a single point in a 2-axis chart, where the x-position of the point indicates a date and the Y-position indicates a move-requirement whose EAD is associated with that date. Obviously, N-dimensional encodings are required for N-ary relations, where the encodings are all attributes of a either a single graphical object. However, to add to the complexity, it is also common to represent different dimensions with multiple objects (e.g. gantt charts with

resources along the y-axis express reservations for each of many activities by using a single interval bar for each activity/interval combination).

Developing an approach to the design of displays of this type required (1) collecting a representative corpus of data examples from both logistical planning and scheduling, as well as from other domains (for generality), (2) identifying the commonalities and differences among these data structures, emphasizing those that are essential for designing graphical displays, (3) understanding the types of graphical techniques which can be combined to represent them, and (4) developing a characterization language with which one can describe data and the expressiveness of graphical technique combinations to enable the representation of one by the other, and (5) developing a theory of graphic design which explicitly defines the syntax and semantics of the components of a diverse range of graphical displays to enable their mapping to complex data structures.

Our accomplishments in this project have included developing a design language which is a system for describing the underlying components of graphics. This language can be viewed as a method for fully specifying a broad range of composite graphics techniques along with the variety of alternative graphic parameters which are available for each technique. The value of this approach is that it allows us to define a new graphic representation very quickly using the language, it provides the ability to construct flexible designs (i.e. a map-like graphic containing graphical objects which can represent supply points, with the flexibility of expressing additional facts about those supply points using many alternative properties: for example, their color, shape, size, links, and/or adjacency to additional textual or graphical objects for each symbol).

The second major advantage to the generality of the flexible design language is that we were able to build an instantiation and rendering module which takes a language plus a data set as input and generates a complete display as output. Thus, any graphic that can be constructed using the design language can be turned into an actual display. Furthermore, the actually mapping process between specific data objects and attributes and graphic properties is automated because of the expressiveness criteria stored with each language. Thus, a user need only specify the data and language to be used and the system uses its knowledge to map attributes to graphic

properties. Finally, the languages are such that they can be composed, so that two graphics can be merged (e.g. aligning multiple charts, integrating a network with a map, merging gauges and textual displays with charts or maps).

The second major area of research has been interactive methods for supporting exploration and manipulation large amounts of multi-dimensional, heterogeneous information. We identified a set of interactive capabilities required to perform many logistical tasks:

- Ability to identify subsets of information relevant to current tasks (e.g., to perform search, to partition available information, to define ranges of information, etc.)
- Ability to control the level of detail with which information is displayed, including the ability to specify computations which are necessary to aggregate, abstract, and summarize information (conversely, to decompose or increase the level of detail with which information is expressed)
- Ability to specify the focus of attention: aspects of the information that are important to display (e.g., specify the attributes, features, or characteristics of information that are relevant to the current task)

Our explorations of these problems have involved a series of design attempts to integrate three types of interactive techniques, which have either been studied in isolation previously or not addressed deeply. Specifically, we have explored three mechanisms: dynamic query, painting and dynamic aggregation.

Dynamic query is a technique for defining the range or scope of information that is desired in a display. It serves the same function as construction of database queries - to select a narrow partition of a large database for current interest. Dynamic query involves the control of multiple sliders, each controlling a different attribute of a data object. As slider values are assigned, the system displays the subset of data objects which satisfy the constraints specified by the query.

Painting is a method which enables subsets of data which are created through either dynamic query or selection from displays representing multiple dimensions to be categorized and represented by color in a painted display. Thus, it is possible to define categories of units

based on their size, movement dates, material requirements, etc (as displayed in one graphic). So large units requiring moderate amounts of fuel to be shipped towards the beginning of a maneuver can be categorized and assigned the color green. This information can then be merged with the location and proximity of the same units on another display, which does not show all this information (e.g. a map). Through methodical use of painting and dynamic query, one can construct through direct manipulation techniques what would otherwise be very complex queries.

Finally, we have been exploring the use of dynamic aggregation and decomposition as methods for navigating through large amounts of information and calculating descriptive statistics of data for analysis purposes. Aggregation consists of two operations: defining a group or set of objects which must be aggregated, and defining a method for abstracting attributes of these objects in the aggregate. For example, a logistician might be viewing the location and fuel requirements (in gallons) for all units in the Army's 18th Corp. Those units that are positioned within a five mile radius of a supply point can be selected through direct manipulation and aggregated to a single object. The fuel quantities for the units can then be abstracted by a user-defined formula (e.g. simple total, mean, etc). The aggregate can then be the focus of new analysis operations, including decomposing it along new dimensions (e.g. by type of unit, by role, by time period in which fuel is needed, etc). Each set of units which is derived from decomposition and aggregation operations can then define painting operations in other displays. For example, it is possible to define quickly a set consisting of all units within a five-mile radius of supply point Charlie, separating out those which use high amounts of type-1 diesel fuel, early in an exercise and then color symbols representing this set in another display which shows the allocation of fuel trucks that can handle that type of fuel on a map.

In summary, our work has been in two areas. The first addresses the problem of visually representing complex data and designing interactively and automatically the composite displays which contain these visual representations. The second area has been concerned with interactive techniques for manipulating the graphic representations of the data. Our current work has not integrated automatic graphics design with interactive techniques, though this is the goal of future research. This work has been published [Roth and Hefley 1993] and will be published in future articles in preparation.

Genetic Algorithm Research:

As discussed in the last review, the objective was the continued development of a genetic algorithm based method to acquire human comprehensible rules for making supply allocation decisions based on examples under various field conditions. Rules can be modeled in the form of If->Then conditional rules, for example: "*if* the terrain is hilly and the weather is overcast and opposing forces are aggressive, *then* request 100 units of ammunition." Finding such rules requires an effective means of search the relationship among the problem elements poses serious difficulties for traditional statistical and symbolic search methods. Our previous research showed the potential of genetic algorithms (GA's) to deal with difficult search through the space of possible feature combinations.

Because field conditions present complex characteristics, effectively modeling problems that included: multiple classifications, interacting relationships among features and noisy (erroneous) data. The previous work led us to the development of COGIN (COverage-based Genetic INduction), a system that uses a novel framework and several new techniques to enable the GA to effectively model decision rules from examples in complex domains. Comparisons with several widely accepted AI induction methods support the superiority of the COGIN approach as a tool for acquiring decision models. This report will describe our progress in developing and
validating COGIN as well as current research efforts to further expand and improve the system.

Results for the Resupply problem. The military's goal was to anticipate (predict) consumption under expected field conditions to provide supplies in a timely fashion. Similar to JIT (just-in-time) manufacturing situations, conditions can change rapidly, the materials can be volatile and perishable, and understocking can have a severe penalty function. The classification rules needed to be comprehensible for use by field supply commanders, little prior domain knowledge was available and applications of traditional operations research techniques had not proven viable. Forecasting resupply requirements is a complex modeling task because of the number of possible parameters (e.g. types of engagement, weather, troop strength, terrain), the potential for non-linear interactions (e.g. individual conditions might suggest increased usage, but when

combined might indicate decreased usage) and the possibility of "noise" or errors in the recorded field data.

Conventional statistical techniques and neural net methods require prior specification of the relevant relationships in the model (or net) and then attempt to find the parameter coefficients (connection weights) necessary to fit the data. Unknown relationships and the inability to interpret coefficients make these approaches undesirable. Symbolic models (such as "if->then" rules) don't require prior specification and can be easily understood, however the number and complexity of the possible combinations make the search task very difficult for conventional symbolic induction methods.

The COGIN solution. The GA has proven to be an effective search method for complex spaces, however conventional application of GA's for symbolic induction from examples can be problematic. COGIN represents both a new methodology as well as a functional system for using the GA to perform inductive search. The fundamental organizing principal of the COGIN search framework is the concept of a *coverage-based constraint* which enables the GA to formulate as simple a model as possible while allowing genetic reproduction to search for the best rules in the model This framework in conjunction with several other innovations has proven very effective as a tool for the acquisition of symbolic decision rules. More generally, the approach provides a new solution to a long standing problem in genetic algorithm research: how to maintain adequate diversity in the population in the context of an optimizing search.

Experimental Results. To evaluate the COGIN system, 50 experimental data sets were created which represented increasingly complex search problems. The primary dimensions were the underlying function, the number of active features, the number of possible classes and the level of noise in the data. Three additional data sets were supplied by the U.S. Army's Human Engineering Labs (HEL) representing the supply scheduling problem. For comparison, the two most widely used symbolic induction systems were also applied. The results were extremely positive, with COGIN outperforming both alternatives across all key complexity dimensions and especially well on the HEL data sets. The results will appear in Machine Learning a primary journal for presenting AI search and learning methodologies [Green and Smith 1992b].

To further evaluate the performance of COGIN a second set of experiments were run using the multiplexor problem which represents a particular class of problems that are difficult because of interactions among the problem features. The experiment compared COGIN against the best known inductive system ID-3 using post-processing. Again the results showed excellent performance despite the very basic configuration of COGIN versus the extensive development of the ID-3 system. The results were presented in July at AAAI-92 in San Jose, California, a major forum for AI techniques and also appear in the proceedings [Green and Smith92a].

Current Research. Much of COGIN's search effectiveness stems from the suitability of the underlying coverage-based concept which was further described and presented at the 1st International Workshop on Learning Classifier Systems, in October in Houston, Texas [Smith and Green 92]. Based on the success of the basic COGIN approach, further development seems extremely promising. Three key areas under evaluation are representational extensions, refinements in the evaluation function, and targeting the search mechanisms. We are currently focusing on methods which allow the representation to adapt to the underlying function variations. Although COGIN outperforms conventional symbolic AI approaches, there are still problem spaces where its predictive accuracy can be improved. Preliminary experiments have shown that COGIN can be made to perform effectively even in problem spaces which are inappropriate for conventional symbolic search. This is especially valuable in modeling supply decision rules since the underlying problem functions can take on many different forms. The ability of COGIN to adapt looks like it offers the most potential.

List of Publications

[Roth & Hefley 1993] Roth, S.F. & Hefley, W.E. "Intelligent Multimedia Presentation Systems: Research and Principles", In Mark Maybury (Ed.) Intelligent Multi-media Interfaces, AAAI Press, Forthcoming.

[Greene & Smith 1992a] Greene, D.P. & S.F. Smith, "COGIN: Symbolic Induction with Genetic Algorithms", In the Proceedings of the Tenth National Conference on Artificial Intelligence.

[Greene & Smith 1992b] Greene, D.P. & S.F. Smith, "Competition-based Induction of Decision Models from Examples" Machine Learning, Forthcoming.

[Smith & Greene 1992] Smith, S.F. and D.P. Greene", "Cooperative Diversity Using Coverage as a Constraint" In Proceedings 1st International Conference on Classifier Systems, Houston, TX, November 1992.

Participating Scientific Personnel

Steven F. Roth, PhD

Stephen F. Smith, PhD

David P. Greene, PhD (obtained PhD, June 1992).

Jade Goldstein